Transcriptomics: Open Access

Khan, Transcriptomics 2018, 6:1
DOI: 10.4172/2329-8936.1000142

Mini Review Open Access

Data Mining – Basics of Bioinformatics

Nida Tabassum Khan*

Department of Biotechnology, Balochistan University of Information Technology Engineering and Management Sciences (BUITEMS), Quetta, Pakistan

Abstract

The world is advancing and developing with the ticking seconds. New technologies, inventions and discoveries have remarkably changed the world as it was before, making lives a lot simpler and easier. From the hunter of gathers society till the scientific society today, many new fields have emerged to combat diseases, make earth a substantially friendly place to live in and fore mostly, making changes for the welfare of mankind. One of such newly developing field is 'Bioinformatics 'which is the use of computers and its software's for acquisition, management and analysis of biological information.

Keywords: Relational database; NoSQL database; Graph database; Flat file database

Introduction

With the advancing discoveries, immense biological data has been generated which needs to be stored. Hence, bioinformatics uses various computational tools to study and analyze the information of biological systems, starting at molecular level. Bioinformatics can be defined as a multi-disciplinary field that incorporates the elements of computer management, database computing, database designing, computer programming, internet networking and molecular biology [1,2]. Bioinformatics can be broadly defined "as the teaching and learning of the use of computer and information technology, along with mathematical and statistical analysis for gathering, storing, analyzing, interpreting, and integrating data to solve biological problems" [3].

Bioinformatics is found to be useful for the following purposes:

- Increasingly biological research relies on information science and hence is managed by computational bioinformatics tools of effective data analysis.

- It has enhanced understanding about the genome structure.

- It has been able to provide storing and sequencing data on databases

- The biological databases have made the complex research work easier.

Now it is used practically in many sub disciplines. Similarly database and knowledge management is also one of the applications of Bioinformatics [4-8].

Databases and knowledge management

This is a very diverse discipline that focuses on several databases and its uses along with the management of the knowledge it provides. The tools used in databases provide us with information in form of knowledge [9] (Figure 1).

Data
Information
Knowledge

Figure 1: Management of the knowledge.

It basically is a combination of two units:

Databases: Databases are basically a computerized collection of data for information retrieval which is shared by many users [10].

Evolution and explanation: The data is processed, organized and stored in databases where it is updated and expanded so that it can be run against applications. It is managed by a computer program, the Database Management System (DBMS). It is a software package that allows the computer to perform database function including storing, modifying and manipulating data [11]. Its history can be stretched back to early 1960's when Charles Bachman created first general Database followed by the introduction of Information Management System (IMS) in 1970 by Codd [12,13] and later in 1980's the relational databases [14]. Today onwards 1990 new data models are used like SQL (Structured query language) databases or cloud databases etc. [15].

Types of databases

- **Relational database:** It is a tabular database in which the data is defined in discrete rows and columns (Table 1) [16].

- **Distributed database:** It is a database in which the data is distributes in many sites [17].

- **Cloud database:** These are the data bases that are accessible on clouds whose primarily function are for online data management [18].

- **NoSQL database:** Non-relational databases [19].

- **Object oriented database:** This is a database that models and create data as objects [20].

- **Graph database:** This provides and arranged data into graphs without the use of rows and columns [21].

- **Flat file database**: This stores database as plain text file and are ideal for small data [22].

- **Biological databases**: This provides information related to genes, nucleic acid, metabolic pathways etc. [23].

Knowledge management: it is the collection of processes that govern the creation, dissemination and utilization of knowledge [24-32].

History and Explanation

Many theorists have contributed to the evaluation of knowledge

***Corresponding author:** Nida Tabassum Khan, Department of Biotechnology, Balochistan University of Information Technology Engineering and Management Sciences (BUITEMS), Quetta, Pakistan, Tel: +92 81 111 717 11; E-mail: 1nidatabassumkhan@yahoo.com

Received December 21, 2017; **Accepted** January 17, 2018; **Published** February 05, 2018

Citation: Khan NT (2018) Data Mining – Basics of Bioinformatics. Transcriptomics 6: 142. doi:10.4172/2329-8936.1000142

Types of Databases	Examples	Knowledge
Nucleotide acid sequences databases (DNA and RNA)	Gen bank	DNA and RNA information [24]
	DDBJ	
	EMBL	
	Rfam	
Protein databases	PSD- Atlas of protein sequences and structures	It provides protein sequence analysis along with protein structure and classification prediction [25]
	PIR-NREF	
	SWISS-PROT	
	Pfam	
	InterPro	
	eMOTIF	
Bibliographic databases	ProClass	Provides literature [26]
Taxonomic databases	NCBI	Gives classification and hierarchical information [27]
	ZooBase-taxonomic databases	
Genomic databases	NCBI	Gives information at gene level [28]
SNP database	dbSNP (Single nucleotide polymorphism Database)	Tells about small scale variations in a gene (insertions or deletions etc. [29]
Phylogenetic databases	RDP (Ribosomal Database Project)	Tells about the ancestor to descendant relationship (i.e. evolutionary relationship) and included phylogenetic ally ordered alignments [30]
Metabolic pathway and enzyme database	KEGG	Provides the information of metabolic pathways and enzymes [31]
	PID	
	HMDB	
	SGMP	

Table 1: Types of databases.

management including Peter Drucker and Peter Senge [33]. The idea was brought up in the beginning of 1970s and further developed with the publications of knowledge management related articles in 1980's [34]. Until now a number of knowledge management firms exist. The knowledge management is the discipline that enables individual, teams or organizations to create and share knowledge collectively so that they can obtain their desired objectives. For instance, research work or case studies that use different tool to produce their collaborated information. For Example Dr. Bob Goldszer; a physician benefits from the knowledge based system by getting any patients tests run against clinical databases to get knowledge of their medical records which can also be compared to others [35,36].

Future Perspectives

Bioinformatics holds great potential in the future by assisting a number of field such as comparative genomics, proteomics, drug discovery, biodenses, microbial genome, system biology, molecular modelling, phylogenetics etc. [37]. It helps in storing the valuable data and provides numerous tools and software packages for manipulation of this data for scientific purpose [38].

Conclusion

Bioinformatics is a new emerging field which uses various computational tools to manipulate and handle biological data. The information tools are very much useful in various disciplines and spheres of life sciences including database and knowledge management.

References

1. Luscombe NM, Greenbaum D, Gerstein M (2001) What is bioinformatics? A proposed definition and overview of the field. Methods Inf Med 40: 346-358.
2. Wang JT, Zaki MJ, Toivonen HT, Shasha D (2005) Introduction to data mining in bioinformatics. Data Mining Bioinformat, Springer London, pp: 3-8.
3. Stein L (2002) Creating a bioinformatics nation. Nature 417: 119-120.
4. Goble C, Stevens R (2008) State of the nation in data integration for bioinformatics. J Biomed Inform 41: 687-693.
5. Baxevanis AD, Bateman A (2006) The importance of biological databases in biological discovery. Curr Protoc Bioinformat 1: 1-1.
6. Kanehisa M, Bork P (2003) Bioinformatics in the post-sequence era. Nature Genet 33: 305-310.
7. Altman RB (1998) A curriculum for bioinformatics: the time is ripe. Bioinformat. Oxford, England. 14: 549-550.
8. Lee TJ, Pouliot Y, Wagner V, Gupta P, Stringer-Calvert DW, et al. (2006) BioWarehouse: A bioinformatics database warehouse toolkit. BMC Bioinformat 7: 170.
9. Liao SH (2003) Knowledge management technologies and applications - Literature review from 1995 to 2002. Expert Sys Appl 25: 155-164.
10. Piateski G, Frawley W, Matheus CJ (1991) Knowledge discovery in databases. MIT press, Cambridge, Massachusetts, USA.
11. Olle TW (2003) Database management system (DBMS), John Wiley and Sons Ltd, NY, USA. pp: 517-520.
12. Bachman CW (1969) Data structure diagrams. ACM Sigmis Database 1: 4-10.
13. Codd EF (1970) A relational model of data for large shared data banks. Commun the ACM 13: 377-387.
14. Suver CA (2000) U.S. Patent No. 6,016,497. Washington, DC: U.S. Patent and Trademark Office.
15. Abiteboul S (1997) Querying semi-structured data. Database Theory-ICDT'97 pp: 1-18.
16. Maier D (1983) The theory of relational databases. Computer Science Press, Rockville, Maryland, USA.
17. Özsu MT, Valduriez P (2011) Principles of distributed database systems. Springer Science & Business Media, Berlin, Germany.
18. Ferretti L, Colajanni M, Marchetti M (2012) Supporting security and consistency for cloud database. Comput Sci 7: 179-193.
19. Han J, Haihong E, Le G, Du J (2011) Survey on NoSQL database. In Pervasive computing and applications (ICPCA), 2011 6th international conference, IEEE. pp: 363-366.
20. Rapley MH, Kennedy JB (1995) Three dimensional interface for an object oriented database. In Interfaces to Database Systems (IDS94), Springer, London. pp: 143-167.
21. Angles R, Gutierrez C (2008) Survey of graph database models. ACM Computing Surveys (CSUR) 40: 1.
22. Fowler GS (1994) cql-A Flat file database query language. In USENIX Winter. pp: 11-21.
23. Stein LD (2003) Integrating biological databases. Nat Rev Genet 4: 337-345.
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: The tool for the unification of biology. Nat Genet 25: 25-29.
25. Hobohm U, Sander C (1995) A sequence property approach to searching protein databases. J Mol Biol 251: 390-399.
26. Korfhage RR (2008) Information storage and retrieval.
27. Beach JH, Pramanik S, Beaman JH (1993) Hierarchic taxonomic databases. Advances in computer methods for systematic biology: Artif Int Databases Comput vision pp: 241-256.
28. Gelbart WM (1998) Databases in genomic research. Science 282: 659-661.
29. Hawken RJ, Barris WC, McWilliam SM, Dalrymple BP (2004) An interactive bovine in silico SNP database (IBISS). Mamm Genome 15: 819-827.
30. Nakhleh L, Miranker D, Barbancon F (2003) Requirements of phylogenetic

databases. In Bioinformatics and Bioengineering, 2003. Proceedings. Third IEEE Symposium, IEEE. pp: 141-148.

31. Bairoch A (2000) The ENZYME database in 2000. Nucleic Acids Res 28: 304-305.

32. Direen HG, Jones MS (2003) Knowledge management in bioinformatics. XML Data Management.

33. Adekola HB (2011) Pragmatic management and the success of Nigerian tertiary in the 21st century.

34. Slavkin HC (2010) Leadership for health care in the 21st century: A personal perspective. J Healthc Leadership 2: 35-41.

35. Davenport TH, Glaser J (2002) Just-in-time delivery comes to knowledge management. Harv Bus Rev 80: 107-111.

36. Bui AA, Morioka C (2010) Information Systems and Architectures. Springer US. Med Imaging Inform 93-137.

37. Kasabov N (2013) Evolving connectionist systems: Methods and applications in bioinformatics, brain study and intelligent machines. Springer Science & Business Media, Berlin, Germany.

38. Nguyen HT, Kreinovich V, Wu B, Xiang G (2012) Applications to bioinformatics. Computing Statistics under Interval and Fuzzy, Springer, Berlin, Heidelberg, Germany. pp: 261-264.